DOCUMENT RESUME

ED 412 224 TM 027 444

AUTHOR	Loup, Karen S.; Ellett, Chad D.
TITLE	Application of the "Personnel Evaluation Standards" to Local District Teacher Evaluation Programs: Analyses of 14 Cases.
PUB DATE	1997-03-00
NOTE	34p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Case Studies; *Educational Policy; Educational Practices; Evaluation Methods; *Professional Development; Program Implementation; Rural Schools; *School Districts; *Standards; State Programs; Suburban Schools; *Teacher Evaluation; Urban Schools
IDENTIFIERS	Connecticut; Louisiana; Personnel Evaluation Standards (Stufflebeam)

ABSTRACT

Local school district teacher evaluation plans, processes, and procedures were studied to determine whether they measure up to the criteria outlined by D. Stufflebeam's "Personnel Evaluation Standards" (1988). The applicability and interpretability of the "Standards" across 14 actual cases from local school districts in Connecticut were analyzed, and these results were compared to those from a previous study of teacher evaluation in Louisiana. The focus was on Connecticut's teacher evaluation cycle as implemented in different districts. Key features of the cycle include appraisal, support, and plans for continued professional growth. Previous study of the System for Teaching and Learning Assessment and Review (STAR) in Louisiana described it as a system that includes a student learning-centered focus that is in contrast to the Connecticut Competency Instrument, a classroom observation system based on the process-product literature of the 1980s. The 14 Connecticut examples included 2 urban, 8 suburban, and 4 rural school districts. Case analysts, 14 Connecticut teachers and administrators, used the categories from the "Standards" (propriety, utility, feasibility, and accuracy) to guide their analyses. Results suggest that Connecticut districts are beginning to consider and implement written policy and procedures for decision making in evaluation, although evaluations as carried out were not fully consistent with the state's declared objectives. However, the analyses do indicate that the "Standards" can be used to outline weaknesses and suggest improvements in local district teacher evaluation programs. Appendixes describe Connecticut's teacher evaluation cycle, the "Personnel Evaluation Standards," and the components of a standard as outlined by D. Stufflebeam (1988). (Contains 3 tables and 11 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 412 224

Application of the *Personnel Evaluation Standards* to Local District Teacher Evaluation Programs: Analyses of 14 Cases

Karen S. Loup
University of Georgia

Chad D. Ellett
Louisiana State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Karen S. Loup
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Paper presented at the annual meeting of the American Educational Research Association
Chicago, Illinois
March, 1997

TM027444

BEST COPY AVAILABLE

Application of the *Personnel Evaluation Standards* to Local District Teacher Evaluation Programs: Analyses of 14 Cases

Introduction

Local district teacher evaluation programs and practices have typically been slow to reflect newer findings from the larger context of research on teaching and learning, from current state-of-the-art knowledge bases on assessment, and from efforts of states who have devoted considerable resources in development of large-scale classroom-based assessment systems to establish instrument validity and reliability and professional credibility of such programs and processes (Ellett & Garland, 1987; Holley, 1979; Loup, Garland, Ellett, & Rugutt, 1996). Key findings of these studies have pointed to several concerns about teacher evaluation instruments, practices, and procedures in use at the local district level. For example: 1) More emphasis was placed on the use of teacher evaluation data for summative (dismissal and remediation) rather than formative (professional development) purposes; 2) Policy bases of local district evaluation systems were somewhat deficient in the areas of establishing performance standards and in implementing comprehensive training programs to train evaluators to make reliable judgements about teaching and learning in classrooms; 3) Few systems allowed for the use of evaluators external to the school district or for the inclusion of peer teachers as assessors; and 4) Local systems had been slow to design procedures to accommodate the potential adverse effects of evaluation context variables on the reliability and credibility of evaluation data and processes.

Concomitant with state and local accountability efforts in implementation of on-the-job teacher evaluation systems, a variety of other national developments have documented the continuing interest in developing sound procedures to evaluate educational personnel (particularly

teachers). Central among these are assessments being developed, piloted and implemented by the National Board of Professional Teaching Standards and the development and federal funding for activities of the Center for Research on Educational Accountability (CREATE) at Western Michigan University. As these fields of evaluation have moved forward, measurement and evaluation professionals have completed considerable work in developing new sets of professional standards for the evaluation of personnel in education. Of particular note here is the 1988 publication of *The Personnel Evaluation Standards* (Stufflebeam, 1988) developed by a national committee of representatives of major professional organizations having a stake in personnel assessment and evaluation in education. This document was designed as a comprehensive resource to be used by those developing, implementing, assessing and improving personnel evaluation programs. The content of the Standards document has wide applicability to personnel evaluation efforts in education and is particularly useful to guide development and judge the quality of programs designed to evaluate teachers and school administrators for the purposes of making professional certification and licensure and employment-related decisions.

Another use of the Standards is to conduct retrospective examinations of evaluation efforts that have been developed and implemented at the state and local district levels in order to broaden understandings of educational reform efforts that view teacher evaluation as a key to school change and improvement. There is only one known study has been completed which has used the Standards as part of a post hoc, historical analysis of large-scale development, implementation, and subsequent demise of statewide teacher evaluation efforts in Louisiana (Ellett,, Wren, Callender, Loup, & Liu, 1996). No known studies completed/published to date that have attempted to use the Standards to analyze local district level teacher evaluation efforts. The study reported here represents an attempt

by university faculty and a group of trained educators to use the Standards as a framework for understanding local, district teacher evaluation systems in use in 14 local school districts in the state of Connecticut. In addition, results are compared with findings in the Ellett, et al. study in an attempt to analyze applicability of the Standards across multiple contexts.

Objectives

The purpose of this paper is to present results of assessing local school district teacher evaluation plans, processes, and procedures used for assessing tenured teachers for the purposes of employment decisions to determine whether they are "measuring up" to criteria outlined by the *Personnel Evaluation Standards*. Of further interest was to examine the applicability and interpretability of the Standards across 14 actual cases from local school districts in Connecticut and comparing these results to findings from a prior study (Ellett et al., 1996) in which the Standards were used to assess a statewide, legislatively mandated system of teacher evaluation for professional and renewable certification of all teachers implemented in the state of Louisiana.

Historical Context of Teacher Evaluation in Connecticut

The State of Connecticut requires on-the-job assessment for all beginning teachers during their first three years of teaching for the purpose of support and issuing of a professional teaching certificate. The Beginning Educator Support and Training (BEST) evaluation program has been developed and implemented by the Connecticut Department of Education (CDE) in local school districts. For beginning teachers obtaining initial certification, the evaluation process is largely controlled by the state. Of pertinent interest in this analysis was the diversity of local evaluation systems developed and controlled by school districts in response to state mandated criteria for a continuous, three-year "Teacher Evaluation Cycle" (Iwanicki, 1990) (Appendix A) for all certified

teachers in Connecticut's schools. Currently, such plans must be periodically approved by the CDE on the basis of suggested, broad guidelines which are drawn from and extend the objectives, processes, and instrumentation included in the BEST program. Guidelines include the assessment and professional support at the local, district level in a three phase, three year cycle: 1) Appraisal; 2) Support; and 3) Continued Professional Growth. The purpose of this cycle is to document a teacher's performance and development over a three-year period through formative and summative processes. Summative evaluation reports are completed during each spring of the cycle.

Key features of this "Teacher Evaluation Cycle" (Iwanicki, 1990) include the following:

1) Appraisal Phase

The Appraisal phase includes intensive evaluation of teachers' performance through a series of formal classroom observations typically conducted by the school principal or appropriate supervisor. Following each formal observation, recommendations are made to strengthen or enhance the teacher's performance and teacher growth is monitored through follow up observations and conferences. Each teacher is appraised every three years, or upon extended completion of all phases of the cycle. In addition, any teacher may be placed in the Appraisal cycle if there is reason to believe that the teacher, in either the Support or Continued Professional Development phases of the cycle, is not performing satisfactorily. The teacher must be informed of this placement in writing (Iwanicki, 1990).

Guiding principles regarding professional responsibilities of teachers for these processes are suggested in Connecticut's Common Core of Learning and specific competencies or proficiencies required of teachers are outlined in the Connecticut Teaching Competencies. The Connecticut Competency Instrument (CCI), a state-developed observation tool (initially developed for use in the

BEST program), grounded in the process-product research of the 1980's, is used for classroom observation and assessment of "generic teaching competencies" or "teacher behaviors" across grade levels. Most local district systems and processes for teacher evaluation and support reflect, for the most part, state guidelines, though a few have extended these ideas to include other instrumentation and processes unique to the district context. For example, in one large suburban school district, a new classroom observation system that included elements beyond the competencies assessed with the CCI was recently adopted.

2) Support Phase

After completing the Appraisal phase, a teacher may enter the Support phase only if a favorable evaluation is received at the conclusion (usually at the end of the school year) of the Appraisal cycle. This cycle allows teachers to assume more responsibility for professional growth by constructing objectives and a plan for monitoring progress toward accomplishment of these objectives. Evaluators provide support and formal or informal classroom observations as necessary.

3) Continued Professional Growth Phase

A teacher enters this phase after completing satisfactory progress toward achievement of objectives at the end of the Support year. This phase is one of self-evaluation and reflection.

Although each phase is expected to be completed in a span of one school year, there are provisions in the cycle to allow more time for teachers to accomplish meaningful objectives in each of the phases. Once the cycle is implemented at the school level, "an evaluator (usually the school principal) works intensely on appraisal with all non tenured teachers in the school and only with about one-third of the tenured teachers (Iwanicki, 1990). Most administrators, supervisors, and department heads have been certified as evaluators for the Connecticut BEST system (for beginning

teachers) and thus, are generally perceived by district officials as able to conduct (via previous training and position responsibilities) credible classroom observations and evaluations for tenured teachers.

Contrasting Case: Teacher Evaluation in Louisiana

A previous analysis using the Personnel Evaluation Standards to assess the development and implementation of a statewide teacher assessment program in Louisiana (Ellett, C., Wren, C., Callender, K., Loup, K., & Liu, X., 1996) was used as a contrasting case in interpretation of the collective results of the fourteen cases in this study. As a result of the passage of a more general legislative mandate for educational reform, The Children First Act (1988) (now amended), Louisiana was the first state in the USA to require on-the-job assessments for the purpose of issuing and/or validating the renewable, professional teaching certificate. In addition, the law provided for salary enhancements for teachers who qualified for and chose to participate in a Model Career Options Program (MCOP).

Historically, in Louisiana, teachers were granted lifetime certification following successful completion of three years of teaching in a local school district. Documentation of success was based on school principals' evaluations which were conducted using a wide variety of locally-developed evaluation systems. In addition to obtaining state certification for " life for continuous service" after this three year period, teachers were typically granted tenure by their respective local school district. Thus, two different decisions (interpreted by most as inseparable), one relative to state certification and the other relative to local tenure, were typically made at one point in time (the three year juncture) in a teacher's career. The new policy guidelines, however, represented a fundamental departure from the historically, time-honored tradition of lifetime certification in Louisiana by

redefining the state level functions in the following ways: 1) The lifetime certificate was replaced with a renewable, professional teaching certificate to be re-validated every five years; and 2) Acquisition of both initial (for beginning teachers) and renewable professional certification (and re-validation) was based upon successful completion of a series of on-the-job classroom assessments in accordance with a statewide developed system (as opposed to the former local system) and program of evaluation.

Thus; much of the confusion in the implementation of the policy involved the distinction between certification as a state decision and tenure as a local, district decision. That distinction was now clearly separated in policy, but for teachers, such practices remained quite contrary to procedures which had been historically accepted as one and the same. Thus, such beliefs affected the perceived validity of any system that might be developed for the state in this response to this policy and had implications for assessment of the evaluation system using the Personnel Evaluation Standards..

An additional component in the Children First Act in Louisiana included the possibility of salary enhancements for teachers who qualified for and chose to participate in a Model Career Options Program (MCOP). In reality, the state chose to implement this option by requiring a "superior" or higher score on components of the assessment system, against recommendations of the system developers. Such a definition of "superior" put undue stress on the assessment system, and caused a great deal of score inflation. Another context consideration of particular note here, is that a series of three year pay increases for **all** teachers in Louisiana was written in the law, and was distributed during the developmental stages (prior to implementation) of the Statewide Teacher Evaluation Program for initial and recertification of teachers in Louisiana.

Thus, teachers received the total number of promised salary increases before the assessment system was implemented and were much more willing to follow union leadership to suspend and amend the policy because of this pre-payment.

The Louisiana State Board of Elementary and Secondary Education (BESE) and the Louisiana Department of Education (LDE) were charged with development and implementation of this comprehensive teacher evaluation program for the purposes of making valid and reliable certification decisions for all teachers as mandated in the policy. This process was a bit unlike the Teacher Evaluation Processes in Connecticut in which the state is responsible for implementation of an initial certification program for beginning teachers (BEST), and local districts are responsible for implementation of evaluation aspects of the Teacher Evaluation Cycle, according to state criteria, but with allowance for individual district contextual concerns.

The specific assessment system used in the Louisiana state program for conducting on-the-job assessments in classrooms was developed by a team of researchers at the state's flagship university through contract agreement with the BESE and LDE. The assessment system, the System for Teaching and learning Assessment and Review (STAR) (Ellett, Loup, & Chauvin, 1990) is a comprehensive, classroom-based observation system designed to assess effective teaching and learning and interactive elements of the learning environment in classrooms. The STAR is considered an example of a new generation of classroom-based assessments that is more comprehensive in focus than earlier classroom observation devices and checklists and existing teacher evaluation instruments. The system is considered an extension of the process-product literature on effective teaching (Brophy, 1986; Porter and Brophy, 1986, etc.) and earlier developments of first generation, large-scale teacher evaluation systems developed for use in other

states (Ellett, Garland, & Logan, 1987) to include a student learning-centered focus grounded in more constructivist views of classroom learning environments. The system stands in contrast to the Connecticut Competency Instrument, which is a classroom observation system based on the process-product literature of the 1980's.

The STAR assessment focus extends earlier developments in significant and important ways. The assessment framework and observation system represents a fundamental departure from more traditional views of evaluation of teacher behaviors to include observing and making inferences about student engagement in learning and the connectedness between effective teaching and learning, as well as the quality of the interpersonal interactions and psychosocial and physical elements of the learning environment. The assessment process represents a holistic view of teaching and learning derived from professional, contextually-based judgments, not from simple rating scale or checklist methodology. As compared to the CCI, the STAR model includes the assessment of critical teaching and learning variables not well represented on past assessment systems such as thinking skills, content emphasis, and learning equity (Ellett, Loup, & Chauvin, 1991). However, consideration of teacher knowledge and training in these complex processes seemed to be somewhat under-examined as the LDE rushed to hastily implement the system in response to policy mandates.

The Louisiana assessment model, like Connecticut's, incorporates an emphasis on-going professional development based upon formative and summative use of assessment data. (Ellett, Loup, & Chauvin, 1991), however, the use of multiple assessors (including a peer), over multiple occasions, is a distinctive feature of the process.

Research and development of the STAR assessment system took place over a two year

pilot period which included research activities related to establishing the validity and reliability of the system in addition to construction and piloting of instrumentation, assessment processes, assessor certification programs, and professional development components of the system as well as conduction of a series of standards-setting activities. These processes, though more comprehensive, were not unlike those conducted for the BEST program in Connecticut and in most states that have implemented statewide teacher assessment programs. Typically, states have had the resources to conduct extensive validity and reliability studies relative to the use of comprehensive, state-wide systems, that local systems have not enjoyed (Loup et al, 1996). Thus, such lack of resources may have effected local systems' abilities to comply with the true spirit and intent of many of the Personnel Evaluation Standards.

Louisiana's implementation of on-the-job assessments for initial and recertification of teachers was initiated in October, 1990. The system was hastily implemented (compared to a thoughtful, slow implementation in Connecticut) in the face of pressure from the governor and mandated implementation deadlines, and what was viewed by the BESE as a few "minor" problems which could be worked out over time.

During this first year of implementation (academic year 1990-91), unaddressed issues and concerns regarding the LDE's capacity for implementation, conflicts among policy stakeholders and special interest groups, confusion created by a series of mis-communications and mis-interpretations between key policy actors, a significant change in the state's governance and subsequent interests in educational issues, the high-stakes nature of assessment decisions, and numerous other contextual conditions combined to create perceived pressure within the educational community regarding the fidelity of implementation of the statewide teacher evaluation program.

As a result of perceived political viability and feasibility issues, the Louisiana Legislature voted in June, 1991, to suspend and re-vamp the ambitious effort only 8 months after its implementation. By contrast, Connecticut's state system was phased in for beginning teachers and initial certification, and state guidelines gave over much authority to local districts for employment decisions. While such processes seemed to be more palatable to teachers, and were implemented without much resistance, the issues in each case (renewable certification in Louisiana, and initial certification and employment in Connecticut) were a bit different. The cases contrasting cases are used in this study as examples of how the Personnel Evaluation Standards might be applied across contexts and purposes for assessing systems for evaluating educators.

Methodology

Sample

Case analyses were conducted for 14 local district teacher evaluation systems in use the state of Connecticut. The sample for the study included two urban, eight suburban, and four rural school districts, considered representative of the urban, suburban, and rural district population distribution across the state.

Procedures and Data Analyses

Case analysts included a team of 14 teachers and school administrators trained in interpretation and use of the *Personnel Evaluation Standards* over a five month period. Each analyst was familiar with the particular district's evaluation system, either having participated as an evaluatee or evaluator. Thus, each analyst provided rich professional perspectives gleaned from their direct experiences in the assessment process and their extensive interactions with practitioners in the everyday life of schools in their districts. Two university faculty members, each with a history of

involvement in research and development in large-scale teacher evaluation efforts, assisted in interpretation of data across districts and comparison of results of Standards analyses with those found in a similar study of a large-scale state teacher evaluation program (Ellett et al., 1996).

The 21 *Personnel Evaluation Standards* are grouped by four categories that "correspond to four basic attributes of sound evaluation" (Stufflebeam, 1988); Propriety Standards (P1-P5), Utility Standards (U1-U5), Feasibility Standards (F1-F3), and Accuracy Standards (A1-A8). A complete listing of each standard relative to the four categories can be found in Appendix B. For each standard, the document includes a set of descriptors which include the following; 1) a definition of the standard in the form of a "should" statement, 2) an explanation of key terms and requirements embodied in the standard, 3) a rationale for inclusion of the standard, 4) a set of guidelines to help evaluators and their audiences meet the requirements, 5) a set of common errors concerning typical problems in conduction of personnel evaluations, 6) one or more illustrative cases which show how the standards might be applied, and 7) a list of supporting documentation to assist the reader in further study in the general realm of the standard (Appendix C).

Each of the categories of standards, Propriety, Utility, Feasibility, and Accuracy, were used in a serial fashion to guide each case analysis. Each Standard was judged according to the set of guidelines and common errors outlined (Appendix B) in the document. In applying each standard, analysts 1) gathered documents and perceptions data, 2) considered the Guidelines and Common Errors included in the Standards document, and 3) made an independent judgement for each of the Guidelines relative to the Standard. Decisions were recorded as "+" (system meets the requirements for the standard) , "-" (system does not meet the requirements for the standard), or NA (standard is not applicable) for each Guideline. Each decision was accompanied by a written rationale. As a

final task, a global judgement was made relative to each district program's compliance, partial compliance, or non-compliance with each of the four Standards categories.

A final review of decisions and rationale was done by university faculty. District analyses results were synthesized and compared with results from the analysis of Louisiana's statewide system to provide perspectives on applicability and interpretability of the Standards across personnel evaluation contexts (i.e., local evaluation for continued employment, and state evaluation for certification).

Results

It is beyond the scope of this paper to discuss every judgement made for each of the 14 cases for each Standard. Results are reported as follows: 1) For the four general attributes categories of the Standards, results are reported as the number of cases judged as in compliance, in partial compliance, and non-compliance with each category (Table 1); 2) For each Standard, the percentage of the maximum possible "+" decisions (excluding N/A decisions) is reported in Table 2; 3) For each standard, the most frequently cited guidelines of non-compliance and common errors are reported in Table 3, and 4) Overall results regarding the application of the Personnel Evaluation Standards to local evaluation systems (in the 14 Connecticut cases) are compared with results of applying the same set of standards to a statewide certification evaluation system developed and implemented in Louisiana.

For the most part, district teacher evaluation policies, procedures, and processes were in compliance with Propriety Standards (Table 1). Ten of fourteen district analyses reported positive decisions for at least 70% of the of standards guidelines across the five standards. Of the remaining four, most at least reported partial compliance (positive decisions ranging from 20-69%).

Collectively these results indicate that most districts have constructed processes and systems with regard for the common welfare of evaluatees and clients and with some forethought as to legal, ethical and due process concerns in personnel evaluation.

For the Utility Standards, although collectively it appears that most districts (10 of 14) were in compliance with the majority of the standards guidelines, most were only in partial compliance with the standard guidelines regarding evaluator credibility. This concern seemed to thematically coincide with decisions made for guidelines for many of the accuracy, propriety, and other utility standards.

For Feasibility standards, most systems were in compliance with guidelines. One system, however, was judged as not complying with many of the guidelines. It should be noted that the Superintendent of this system had recently directed all schools to implement a new process and instrumentation for classroom observation. The new system was viewed by teachers and administrators as having been hastily implemented, too cumbersome, and developed with limited teacher input. Thus, negative perceptions of the system and processes may indeed have contributed to the judgement of its limited political viability.

District systems tended to be (to a lesser degree than in other areas) in compliance with Accuracy Standards, particularly as they related to establishment of the validity and reliabilities of their instruments, systems and processes. Districts seemed to rely heavily on their delegated authority for evaluation as a substitute for gathering and analyzing data on the validities and reliabilities of their evaluation measures. Again, in the case of one district judged as being out of compliance with many of the guidelines, it was obvious that a system was being implemented without much forethought to validity and reliability issues and their accompanying legal and ethical implications.

Table 1: Compliance Judgements for Attributes of the Personnel Evaluation Standards for 14 Cases

Standard Attribute	Compliance	Partial Compliance	Non-Compliance
PROPRIETY	10*	4	--
UTILITY	10	4	--
FEASIBILITY	12	1	1
ACCURACY	9	4	1

* Number of Cases

For each of the twenty-one standards, the percentage of the maximum possible "+" decisions regarding the system's meeting of each of the guidelines is reported in Table 2 for the 14 cases collectively.

Table 2: Percentage of the Maximum Possible Acceptable Decisions for Standards Guidelines for 14 Cases

Standard		% of Max Possible
P: Propriety Standards		
P1	Service Orientation (13)*	90
P2	Formal Evaluation Guidelines (17)	75
P3	Conflict of Interest (9)	82
P4	Access to Personnel Evaluation Reports (15)	46
P5	Interactions With Evaluatees (7)	70
U: Utility Standards		
U1	Constructive Orientation (11)	85
U2	Defined Uses (6)	82
U3	Evaluator Credibility (13)	38
U4	Functional Reporting (8)	79
U5	Follow-Up and Impact (15)	79
F: Feasibility Standards		
F1	Practical Procedures (10)	82
F2	Political Viability (7)	93
F3	Fiscal Viability (6)	95

BEST COPY AVAILABLE

Table 2: continued

A: Accuracy Standards		
A1	Defined Role (7)	82
A2	Work Environment (2)	76
A3	Documentation of Procedures (3)	80
A4	Valid Measurement (10)	36
A5	Reliable Measurement (7)	52
A6	Systematic Data Control (15)	81
A7	Bias Control (9)	72
A8	Monitoring Evaluation Systems (9)	60

* Number of Guidelines Defining each Standard

The lowest percentages of acceptable decisions across school districts were for the Propriety Standard of Access to Personnel Evaluation Reports (P4) (46%), the Utility Standard of Functional Reporting (U4), and the Accuracy Standards of Valid Measurement (A4) (36%),Reliable Measurement (A5) (52%). Most districts systems were judged as being largely in compliance with the Propriety Standard of Service Orientation (P1) (95%) and the Feasibility Standards of Political Viability (F2) (93%) and Fiscal Viability (F3) (95%).

The most frequently cited lack of adherence with guidelines and common errors were also noted for each standard and can be found in Table 3.

Table 3: Most Commonly Cited Lack of Adherence with Guidelines and Common Errors for 14 Cases

Standard	Guideline	Common Error
P: Propriety Standards		
P1 Service Orientation	L. Enforce prescribed standards consistently.	B. Being overprotective of, or oblivious to, incompetent or harmful personnel. C. Failing to invest resources in the training of evaluators.
P2 Formal Evaluation Guidelines	H. Involve board and staff in development and periodic review of policies & guidelines. K. Assure consistent enforcement of the written evaluation guidelines.	C. Being insensitive to the fact that in addition to the evaluation process being fair and equitable, it must be perceived as such by all concerned.

Table 3: continued

		P. Establish a process for periodic review And revision of evaluation procedures and guidelines.	
P3	Conflict of Interest		
		D. Exercise control of conflict of interest at every level of examination & judgement. E. Employ evaluation procedures requiring comparison of multiple sources of information to discover any tainted evidence.	B. Failing to define the conditions under which the evaluation will be considered valid.
P4	Access to Personnel Evaluation Reports		
		H. Notify an evaluatee in writing when the the institution has added sensitive or possibly controversial information or documents to personnel files. I. Provide each evaluatee continuing opportunity to review his/her file, append (within time restraints) written comments, and request a copy of any item contained in the file. J. Specify in writing that, subject to statutory limitations, access, retrieval, and release of evaluation reports should be limited to persons with a legitimate need to know. K. Make arrangements for secure storage of evaluation reports and other evaluation records.	A. Failing to distinguish between a legitimate need to have access to an evaluation report and an expressed need based on curiosity or some other inappropriate purpose. B. Granting access to persons of some standing but with no legitimate need to see a file or a report. C. Failing to give timely notice when new information is added to a file.
P5	Interactions With Evaluatees		
		D. Monitor the effectiveness of the evaluation regularly through systematic collection of process feedback from evaluatees. E. Schedule evaluation activities well in advance and stick to the schedule.	A. Failing to create the conditions for timely and constructive interaction between evaluatees and evaluators.
U: Utility Standards			
U1	Constructive Orientation		
		F. Provide timely evaluation feedback. J. Use evaluations to allocate resources for for improving performance, and provide resources and support for that purpose.	A. Assuming that personnel evaluation practices, procedures, and objectives are self-explanatory and acceptable. C. Failing to assess whether the educator is provided with sufficient resources and support to do the job.
U2	Defined Uses		
		B. Invite the evaluatees to help determine evaluation goals, uses, forms, methods, and audiences. F. Monitor the evaluation process to ensure tight connections between the collected information, intended uses, and actual uses.	A. Assuming that all users have identical or similar needs that will be met by the same type of information.

Table 3: continued

U3	Evaluator Credibility	A. Assign evaluation roles to educators with appropriate professional training and skills, professionalism & sensitivity, and who understand evaluation tasks and roles of personnel to be evaluated. B. Ensure that evaluators of classroom practice understand effective teaching techniques and principles of learning psychology. C. Train administrators, board members faculty, and evaluation specialists to be effective in roles in the evaluation system. D. Train those who will serve as evaluators in principles of sound personnel evaluation, performance appraisal techniques, methods for motivating faculties, conflict management, and the law as it applies to education personnel evaluation. F. When feasible, engage an evaluation team, rather than a single administrator, to enhance credibility and validity. G. Provide evaluators with support personnel or services to assist in collecting and analyzing needed information when the tasks exceed their professional training and expertise. M. Evaluate the work of each evaluator periodically.	A. Defending & sustaining a weak or ineffective personnel evaluation system. B. Failing to provide a climate of support and growth for evaluatees. C. Failing to plan and prepare carefully for observations & other evaluation activities. F. Basing evaluation conclusions on preconceptions rather than valid information. G. Failing to provide evaluatees with recommendations or services they can use to overcome identified deficiencies. H. Discussing evaluation results with persons who have no professional reason to know them.
U4	Functional Reporting	A. Write the report immediately following the observation.	C. Overemphasizing the importance of strengths or weaknesses through use of excessive, misleading detail. E. Presenting the report as reflecting consensus though contributors failed to agree on recommendations.
U5	Follow-Up and Impact	F. Arrange follow-up conferences between evaluatee & appropriate support personnel. J. Keep written record of instances in which evaluatee did not act on evaluator recommendations.	E. Failing to keep superiors informed of serious performance problems. F. Failing to use evaluation information appropriately in personnel decisions.
F: Feasibility Standards			
F1	Practical Procedures	I. Review procedures periodically to to assess how they could be strengthened. H. Identify and assess published evaluation procedures as step toward improving local procedures	B. Adding or omitting procedures arbitrarily while the evaluation is in progress. D. Unduly stressing practicality over accuracy.

Table 3: continued

F2	Political Viability		
		C. Provide sufficient time & opportunity for concerned individuals and groups to develop, review, and revise personnel evaluation policies & procedures. E. Review personnel policies periodically.	A. Assuming that cooperation happens automatically. G. Assuming that rationales, norms or rules of tenure legally shield any educator from accountability for performance.
F3	Fiscal Viability		
		C. Estimate the personnel time required to conduct each type of personnel evaluation, and use the estimates to allocate staff time accordingly.	B. Failure to allocate human and fiscal resources to maintain the system.
A: Accuracy Standards			
A1	Defined Role		
		C. Define duties that reflect the needs of students, constituency, and employing institution.	B. Evaluating performance on aspects of position that exist in written documents such as formal descriptions, even though unrelated to actual position.
A2	Work Environment		
		A. Identify and record all variables that might affect the work environment.	D. Failing to state how contextual variables will be used in interpreting assessment data.
A3	Documentation of Procedures		
		B. Provide all evaluatees and other users with feedback forms on which to suggest improvements in the evaluation system.	B. Failing to document exceptions to the design procedure.
A4	Valid Measurement		
		B. Ensure that plans for full implementation of measurement procedures are based on careful review of intended uses. C. Involve those affected by the system in determining its purposes, processes, assessment criteria, instruments, and in assessing its validity. F. Field test the measurement procedures using appropriate validation techniques. G. Insure the validity of any measurement procedures that disproportionately affect members of any identifiable group. H. Make the results of the validation process public and describe results in terms that are understandable and meaningful to evaluatees and other interested individuals & groups. I. Report validity results openly and completely and include descriptions of what inferences are supported by validity evidence & what precautions must be taken in using them for decision making.	A. Concentrating on parts of the job that are easiest to assess. C. Using a measurement procedure for multiple purposes when it is valid for only one. F. Assuming that a procedure is valid solely because it "seems reasonable" or is common practice. H. Failing to document the validation process.

Table 3: continued

		J. Encourage sufficient flexibility in negotiated contracts, legislation, and board policy so that evaluation instrumentation and measurement procedures can be improved to enhance validity over time.	
A5	Reliable Measurement		
		A. Acquire evidence for all types of reliability that are relevant to intended uses of instrumentation before using it in evaluation. B. Check for inconsistency in interpretations of measurement results. C. Estimate and report reliability of instruments for the particular situation. D. Train observers to apply the criteria consistently and objectively.	A. Confusing reliability with validity...necessary but not sufficient condition for validity. B. Assuming that the reliability of a procedure is the same for different groups and situations. C. Assuming that published reliability estimates are necessarily applicable to intended use.
A6	Systematic Data Control		
		M. Maintain complete and well-documented records on all evaluation follow-up.	B. Failing to maintain an "audit trail" for the data.
A7	Bias Control		
		E. Obtain data and judgements from multiple sources and preserve independent data and judgements for possibility of independent review.	A. Continuing to use evaluators whose biases have become evident.
A8	Monitoring Evaluation Systems		
		A. Investigate whether the evaluation system is having a positive effect on the quantity and quality of educational outputs. B. Budget sufficient resources and personnel time to review the evaluation processes regularly. C. Identify parts of the system that require frequent review or close monitoring. E. Periodically survey staff to obtain recommendations. G. Review policies and plans against the Standards and other relevant sources. H. Train evaluatees, evaluators, and others in using the Standards to evaluate the system. U. Check that the validity and reliability information is current and accurate.	A. Assuming that a well-developed and carefully implemented performance evaluation system will continue to operate as well in succeeding years as it does the first year. B. Failing to train agency staff for their roles in personnel evaluation beyond the initiation year. E. Revising the evaluation system without consulting the users.

Most of the problems encountered in judging district systems as in compliance with several of the Propriety standards were related to developing processes and procedures for storage, access, and additions to personnel records. Without such forethought and training for responsible staff, access problems are likely to occur. In one case, for example, the curriculum supervisor for the

system was granted access to all evaluation records for the purpose of making a routine decision regarding attendance at a national meeting, hardly a legitimate purpose for viewing such data.

By contrast, in the Louisiana statewide case, compliance with the Propriety Standards (particularly the P4 Standard) was relatively strong as the system was designed in response to legalities inherent in making high-stakes certification, rather than local employment decisions. Access to personnel evaluation records was heavily controlled by the state, and districts were made aware of ramifications for lack of compliance.

Another notable area of difficulty among the 14 district cases was for the Utility Standard of Evaluator Credibility (U3). In most districts, the principal of the school tended to be the sole evaluator of teachers. Most did not have comprehensive training programs at the local level, and the inference was made by the district that evaluators who had been through the state-conducted BEST evaluator training program (with periodic update sessions encouraged) were competent to conduct local evaluations. It should be recognized that many of the districts used the Connecticut Competency Instrument (CCI) (also used in the BEST program) as the basis for classroom observation evaluations. However, many of the districts had developed their own modifications to the CCI and/or other systems and instruments for which assessor training involved little more than "informing". Thus, these districts seem to consider positional authority along with completion of BEST assessor training as the basis for establishing evaluator credibility. Although districts believed that principals indeed possessed and practiced appropriate classroom observation and conferencing skills, many teachers questioned evaluator credibility in that principals were not considered knowledgeable of all subject matter areas and appropriate pedagogical methods.

By contrast, in the Louisiana Case, the certification system was judged as receiving rather high marks regarding evaluator credibility. Again, this compliance might have been a result of the

fact that the program had been through two current years of research and development in which considerable resources had been allocated for assessor training and certification. The evaluator training involved an intensive seven-day program, which included multiple proficiency requirements, a field component, and follow-up classroom assessments. It was of utmost interest to the state to establish evaluator credibility and proficiency as classroom observation data would be used for high-stakes decisions such as renewal of the teaching certificate, and career incentives. In Connecticut, however, use of the observation data for employment decisions was not perceived to be a high-stakes activity. Teachers perceived the process as somewhat ritualistic (Loup et al., 1996) and, even though most did not perceive evaluators as credible, they were willing to accept the process as a simple requirement for annual review with few ramifications for their everyday lives in classrooms.

For Feasibility standards, most systems were in at least partial compliance with guidelines for political viability as opposed to the statewide results in Louisiana which indicated that political viability was a major area of non-compliance. Again, these compliance decisions in Connecticut may have been a result of the low-stakes nature of the decisions and the nature of the political culture of the state. In Connecticut, teacher unions have collective bargaining agreements with school districts. The fact that the organizations sanctioned the local processes rendered the systems to some degree politically feasible and viable. In addition, the state of Connecticut has devoted considerable resources to the hiring of State Department of Education Personnel with a plethora of expertise in research and development. Thus, State Department directives are seen in a much more credible light than in many states. Thus, given that most systems were seen as in compliance with state mandates, and the fact that most used the CCI as the primary classroom observation instrument, it was generally assumed that local systems were appropriate and feasible.

By contrast, the Louisiana Teacher Evaluation Program was developed and implemented for the purposes of making high-stakes certification decisions for all teachers on a renewable basis. Implementation by the State Department of Education was characterized by hastily developed and poorly thought through policies, lack of monitoring and resources, compounded by a negative political agenda of teacher unions and the media aimed at discrediting the program. Such a context for implementation essentially resulted in perceptions that the program was not politically viable.

District systems in Connecticut were judged to be low in compliance with Accuracy Standards, particularly as they related to establishment of the validity (A4) and reliabilities (A5) of their instruments, systems and processes, and system monitoring and review (A8) . Again, in contrast to the results of the Louisiana case in which this set of standards was judged as a strong area of compliance, districts seemed to rely heavily on their delegated authority for evaluation as a substitute for gathering and analyzing data on the validities and reliabilities of their evaluation measures. Though most local systems were not perceived as valid and reliable for actually assessing classroom teaching and learning, most teachers were willing to accept them as valid because, as the culture dictated, such systems were accepted as ritualistic rather than professional, the stakes were not too high, and most teachers passed. In one particular case, however, a local superintendent had initiated a new system, and participants generally perceived it as invalid for assessing classroom teaching and learning. There were no validity and reliability studies planned, and assessors (as well as teachers) were not being trained in the new system other than simply at an informational level. In another system, the third phase of the Teacher Evaluation Cycle (reflection and self-assessment) was seen by most as a "gift" year for compliance with the Appraisal and Professional components of the cycle.

Concerns for monitoring constituted a theme that permeated both the Connecticut and Louisiana case analyses. There seemed to be lack of attention to detail in this arena as a result of either lack of resources, knowledge, or realization of the relationship between monitoring and perceptions of system credibility in both instances. Such insufficient attention often led to opportunities for abuses and misuses of the systems.

Discussion and Implications

Results of this study are important from a variety of perspectives. First, the analysis task demonstrated that the Personnel Evaluation Standards could be used to delineate weaknesses and suggest improvements in local district teacher evaluation programs. The exercise of using the Standards as demonstrated in this study, is particularly recommended for those seeking to develop similar insights and common understandings about the technical, human, political and social factors that influence the quality of development, implementation and evaluation of local, district personnel evaluation efforts.

As found in the Louisiana comparison study, the Standards appear to be an excellent source of new learning and professional debate and discussion. Secondly, when results of this local district study were compared to the larger statewide study, it became apparent that the Standards take on multiple meanings from multiple perspectives and should be understood in this vein. From that of a technical measurement perspective, as indicated in results of the Louisiana study, there was sufficient evidence for the psychometric validity of the evaluation measure to strongly support this validity concern relative to the Accuracy Standards. However, when combined with the fact that teachers perceived the system and measure to be invalid, evidence of psychometric validity alone seemed to be insufficient for judging compliance with this standard. Similarly, when considering results across the 14 local cases in this study, compliance with political viability seemed to be

somewhat of a tradeoff for gathering evidence of psychometric validity.

Results of this study when combined with studies of local school district evaluation systems (Loup et al., 1996) seem to suggest that local school districts apparently have at least begun to consider and implement written policy and procedures for decision making in evaluation processes. However, case analysts reported that few district systems included guidelines for monitoring the consistency with which teacher evaluation processes are implemented, and that evaluations were carried out in a ritualistic manner that resulted little professional growth and development, quite inconsistent with Connecticut's philosophical guidelines for the evaluation and growth cycle, and relatively inconsistent with recommendations and national criteria and guidelines outlined in the Personnel Evaluation Standards (1988) for assessing systems for evaluating educators.

From a reform and school improvement perspective, as Cuban, 1990 indicates, reforms tend to be cyclical and faddish, implemented in a cursory manner with lack of consideration of time needed to implement and evaluate in meaningful fashions. Perhaps, newer ideas in development of more professional models of assessment and evaluation, as evidenced in the work of the National Board for Professional Teaching Standards, the work of centers like CREATE, and dialogue surrounding the *Personnel Evaluation Standards* will serve as springboards for planning, developing, implementing, and evaluating local and state systems such as those described here.

References

Cuban, L. (1990). Reforming again, again, and again. Educational Researcher, 19(1), 3-13.

Brophy, J. (1986, October). Teacher influences on student achievement. American Psychologist, 1069-1077.

Ellett, C.D., & Garland, J. (1987). Teacher evaluation practices in our largest school districts: Are they measuring up to 'state-of-the-art' systems? Journal of Personnel Evaluation in Education. 1(1), 69-92.

Ellett, C., Loup, K., & Chauvin, S. (1990). The System for Teaching and learning Assessment and Review (STAR). Assessment manual. Baton Rouge, LA: Louisiana Teaching Internship and Statewide Teacher Evaluation Projects, College of Education, Louisiana State University.

Ellett, C., Loup, K., & Chauvin, S. (1991). Development, validity and reliability of a new generation of assessments of effective teaching and learning: Future directions for the study of learning environments. Journal of Classroom Interaction, 26(2), 25-39.

Ellett, C., Wren, C., Callender, K., Loup, K., & Liu, X. (1996). Looking backwards with the Personnel Evaluation Standards: An analysis of the development and implementation of a statewide teacher assessment program. Studies in Educational Evaluation, 22(1), 79-113.

Holley, F. (1979). Teacher testing, teacher evaluation and research on teaching: Converging fields? Paper presented at the annual meeting of the American Educational Research, San Francisco, CA.

Iwanicki, E. (1990). Teacher evaluation for school improvement. In J. Millman & L. Darling-Hammond (Eds.), The new handbook of teacher evaluation. Newbury Park, CA: Sage.

Loup, K., Garland, J., Ellett, C., & Rugutt, J. (1996). Ten years later: Findings from a replication of a study of teacher evaluation practices in our 100 largest school districts. Journal of Personnel Evaluation in Education, 10(4), 203-226.

Porter, A.C., & Brophy, J. (1988, May). Synthesis of research on good teaching: Insights from the work of the institute for research on teaching. Educational Leadership, 45(8), 74-85.

Stufflebeam, D. (1988). The personnel evaluation standards: How to assess systems for evaluating educators. Newbury Park, CA: Sage.

APPENDIX A

The Teacher Evaluation Cycle in Connecticut

Iwanicki (1990)

An Overview of the Teacher Evaluation and Professional Growth Cycle

Year 1 - Appraisal

Focus: To conduct a thorough appraisal of teacher performance in light of the indicators of effective teaching.

Events: a) Fall conference to discuss and initiate the appraisal process
b) Minimum of three classroom observations, each with follow-up conferences resulting in a written **Classroom Observation Report**
c) Spring evaluation conference to discuss the teacher's **Appraisal Report** and to develop a Professional **Growth Plan** for the next two years. This plan is based on objectives which can focus on
1) Strengthening performance with respect to the indicators of effective teaching
2) Professional growth initiatives
3) School improvement targets

Orientation: This is a collegial but accountability-oriented process where the evaluator assumes leadership for evaluating teaching. The teacher and the evaluator use this information as they work collaboratively to develop a two-year **Professional Growth Plan** to strengthen or enhance the teacher's performance.

Year 2 - Support

Focus: To support the teacher as work begins on the objectives which serve as the basis of his/her Professional Growth Plan and to monitor progress in this regard

Events: a) Fall conference with supervisor to review objectives and to decide how the teacher's progress will be supported and monitored
b) At least two conferences to support and to monitor the teacher's progress toward objectives
c) Formal classroom observations and informal class visits as necessary
d) Spring conference to complete the teacher's **Spring Progress Report on Objectives**

Orientation: This is a collegial partnership, where the supervisor supports and guides the teacher's efforts to achieve the objectives which serve as the basis of his/her Professional Growth Plan.

Year 3 - Continued Professional Growth

Focus: To provide the teacher with the opportunity to
a) pursue what needs to be done to achieve the objectives which serve as the basis of his/her Professional Growth Plan
b) conduct a self-evaluation and to reflect on where he/she is going professionally

Events: a) Fall conference with supervisor to review what the teacher needs to do to achieve his/her objectives, to discuss what strategies the teacher might use to conduct a self-evaluation, and to reflect upon his/her performance
b) Interim evaluation conferences as necessary
c) Formal classroom observations and informal class visits as necessary
d) In spring the teacher completes the **Final Evaluation Report**, which is forwarded to the evaluator. This report includes a self-assessment of
1) the extent to which the teacher's objectives have been achieved and
2) those indicators of effective teaching which will provide the focus for his/her future professional development

Orientation: This is a reflective process, where the teacher assumes a more direct role in evaluating his/her performance and in setting a direction for future professional development.

* Adapted from Iwanicki, 1990

BEST COPY AVAILABLE

APPENDIX B

The Personnel Evaluation Standards

Stufflebeam (1988)

THE PERSONNEL EVALUATION STANDARDS

The Joint Committee on Standards for Educational Evaluation

P1 ***PROPRIETY STANDARDS***

- P1 Service Orientation
- P2 Formal Evaluation Guidelines
- P3 Conflict of Interest
- P4 Access to Personnel Evaluation Reports
- P5 Interactions with Evaluatees

U ***UTILITY STANDARDS***

- U1 Constructive Orientation
- U2 Defined Uses
- U3 Evaluator Credibility
- U4 Functional Reporting
- U5 Follow-Up and Impact

F ***FEASIBILITY STANDARDS***

- F1 Practical Procedures
- F2 Political Viability
- F3 Fiscal Viability

A ***ACCURACY STANDARDS***

- A1 Defined Role
- A2 Work Environment
- A3 Documentation of Procedures
- A4 Valid Measurement
- A5 Reliable Measurement
- A6 Systematic Data Control
- A7 Bias Control
- A8 Monitoring Evaluation Systems

APPENDIX C

Components of a Standard

The Personnel Evaluation Standards

Stufflebeam (1988)

Components of a Standard

Descriptor:	For example, Service Orientation
Standard:	A definition of the standard in the form of a "should" statement; i.e., the Service Orientation standard is stated as follows: "Evaluations of educators should promote sound education principles, fulfillment of institutional missions, and effective performance of job responsibilities, so that the educational needs of students, community, and society are met."
Explanation:	A conceptual introductory statement that defines key terms in the standard and describes the essence of the requirements embodied in the standard.
Rationale:	A generalized argument for the inclusion of the standard.
Guidelines:	A list of procedural suggestions intended to help evaluators and their audiences to meet the requirements of the evaluation standard. These are strategies to avoid mistakes in applying the standard. The guidelines should not be considered exhaustive or mandatory. Rather, they are procedures to consider and to follow when the evaluator judges them to be potentially helpful and feasible.
Common Errors:	Warnings about typical problems in conducting personnel evaluations, as well as possible negative side effects resulting from taking one standard too seriously without considering its impact on other standards.
Illustrative Case:	One or more illustrations. These show how the standard might be applied, including the description of a certain setting, a situation in which the standard is not met, an analysis of the attending problems, and a discussion of corrective actions that would result in meeting the standard. The corrective actions discussed are only illustrative and are not intended to encompass all possible corrections. As much as possible, the illustrative cases have been based on actual evaluations. Also, each case is slanted directly to highlight the salient points in the particular standard, rather than to encompass points across all of the standards.
Supporting Documentation:	Selected references listed at the end of each standard to assist the reader in further study in the general realm of the standard.

* Adapted from Stufflebeam, 1988